

# Catastrophic forgetting and mode collapse in GANs

Year: 2020 | Citations: 248 | Authors: Hoang Thanh-Tung, T. Tran

---

## Abstract

In this paper, we show that Generative Adversarial Networks (GANs) suffer from catastrophic forgetting even when they are trained to approximate a single target distribution. We show that GAN training is a continual learning problem in which the sequence of changing model distributions is the sequence of tasks to the discriminator. The level of mismatch between tasks in the sequence determines the level of forgetting. Catastrophic forgetting is interrelated to mode collapse and can make the training of GANs non-convergent. We investigate the landscape of the discriminator's output in different variants of GANs and find that when a GAN converges to a good equilibrium, real training datapoints are wide local maxima of the discriminator. We empirically show the relationship between the sharpness of local maxima and mode collapse and generalization in GANs. We show how catastrophic forgetting prevents the discriminator from making real datapoints local maxima, and thus causes non-convergence. Finally, we study methods for preventing catastrophic forgetting in GANs.